UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/044,782	01/11/2002	Andreas Arning	DE920000057US1	8377

25259 7590 05/27/2004

IBM CORPORATION
3039 CORNWALLIS RD.
DEPT. T81 / B503, PO BOX 12195
REASEARCH TRIANGLE PARK, NC 27709

EXAMINER	
LY, ANH	

ART UNIT	PAPER NUMBER
2172	5

DATE MAILED: 05/27/2004

Please find below and/or attached an Office communication concerning this application or proceeding.

Office Action Summary	Application No.	Applicant(s)	
	10/044,782	ARNING ET AL.	
	Examiner	Art Unit	
	Anh Ly	2172	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If the period for reply specified above is less than thirty (30) days, a reply within the statutory minimum of thirty (30) days will be considered timely.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on *11 January 2002*.

2a)☐ This action is **FINAL**. 2b)☒ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) *1-13* is/are pending in the application.

4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) *1-13* is/are rejected.

7)☐ Claim(s) _____ is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☐ The specification is objected to by the Examiner.

10)☐ The drawing(s) filed on _____ is/are: a)☐ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☒ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a)☐ All b)☒ Some * c)☐ None of:

1.☒ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☐ Notice of References Cited (PTO-892)
2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
3) ☐ Information Disclosure Statement(s) (PTO-1449 or PTO/SB/08) Paper No(s)/Mail Date _____.
4) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____ .
5) ☐ Notice of Informal Patent Application (PTO-152)
6) ☐ Other: _____.

DETAILED ACTION

1. This Office Action is response to Applicants' communications filed on 01/11/2002.

2. Claims 1-13 are pending in this application.

Claim Rejections - 35 USC § 103

3. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

4. This application currently names joint inventors. In considering patentability of the claims under 35 U.S.C. 103(a), the examiner presumes that the subject matter of the various claims was commonly owned at the time any inventions covered therein were made absent any evidence to the contrary. Applicant is advised of the obligation under 37 CFR 1.56 to point out the inventor and invention dates of each claim that was not commonly owned at the time a later invention was made in order for the examiner to

consider the applicability of 35 U.S.C. 103(c) and potential 35 U.S.C. 102(e), (f) or (g) prior art under 35 U.S.C. 103(a).

5. Claims 1-13 are rejected under 35 U.S.C. 103(a) as being unpatentable over US Patent No. 6,507,840 issued to Ioannidis et al. (hereinafter Ioannidis) in view of Patent No. 6,636,862 issued to Lundahl (hereinafter Lundahl).

With respect to claim 1, Ioannidis teaches a) determining a foreground frequency of a bucket within a first cluster (using histogram technique to determine the bucket and its frequencies for the data distribution sets: col. 8, lines 62-67 and col. 9, lines 1-30).

b) determining a background frequency of the bucket with respect to all of the clusters (ranges of attribute values into buckets: col. col. 10, lines 1-48);

c) comparing the foreground and background frequencies (comparing the data distribution sets: col. 6, lines 52-67).

Ioannidis teaches using bucket histogram technique for data clustering, the distance between of two multisets, in parallel processing database systems (col. 5, lines 25-50 and col. 6, lines 32-50), each bucket is assuming that the values that fall within the range of a bucket (col. 5, 52-67; also see col. 2, lines 16-65) and the frequencies representing sets of two data distributions (distance measurement based on various distribution moments) and comparing the data distribution sets. Ioannidis does not explicitly teach d) determining a quality index based on the comparison.

However, Lundahl teaches the index for a given data clustering (col. 13, lines 43-67 and col. 14, lines 1-8).

Therefore, it would have been obvious to a person of ordinary skill in the art at the time the invention was made to combine the teachings of Ioannidis with the teachings of Lundahl so as to compare the value of the index of a given data clustering and proving the best index value is selected (col. col. 15, lines 1-25; also see fig. 3). The motivation being to have a database for storing the data in which provides the clustering or partitioning advantage for a wider range of queries and used the bucket histogram technique to compare the distance of data distribution sets and the results from the clustering algorithms in a parallel processing system environment.

With respect to claim 2, Ioannidis teaches wherein said comparing step further comprises subtracting the relative foreground and background frequencies (during computing the distance of data distribution of bucket sets: col. 7, lines 15-67).

With respect to claim 3, Ioannidis discloses a method as discussed in claim 1.

Ioannidis teaches using bucket histogram technique for data clustering, the distance between of two multisets, in parallel processing database systems (col. 5, lines 25-50 and col. 6, lines 32-50), each bucket is assuming that the values that fall within the range of a bucket (col. 5, 52-67; also see col. 2, lines 16-65) and the frequencies representing sets of two data distributions (distance measurement based on various distribution moments) and comparing the data distribution sets. Ioannidis does not explicitly teach d) determining a quality index based on the comparison. Ioannidis does not explicitly teach squaring the result of the comparison.

However, Lundahl teaches the sums of squares matrices for each cluster (col. 27, lines 18-67).

Therefore, it would have been obvious to a person of ordinary skill in the art at the time the invention was made to combine the teachings of Ioannidis with the teachings of Lundahl so as to have sum of squares of the result of the cluster to be compared and compare the multi-dimensional qualities data structure having multiple partitions and clusters that can be constructed for the same data. The motivation being to have a database for storing the data in which provides the clustering or partitioning advantage for a wider range of queries and used the bucket histogram technique to compare the distance of data distribution sets and the results from the clustering algorithms in a parallel processing system environment.

With respect to claim 4, Ioannidis discloses a method as discussed in claim 1. And Ioannidis teaches updating up-to-date the database for processing operation (col. 3, lines 1-18).

Ioannidis teaches using bucket histogram technique for data clustering, the distance between of two multisets, in parallel processing database systems (col. 5, lines 25-50 and col. 6, lines 32-50), each bucket is assuming that the values that fall within the range of a bucket (col. 5, 52-67; also see col. 2, lines 16-65) and the frequencies representing sets of two data distributions (distance measurement based on various distribution moments) and comparing the data distribution sets. Ioannidis does not explicitly teach e) determining an optimal number of clusters; and f) comparing the optimal number of clusters to the actual number of clusters.

However, Lundahl teaches the optimal of the number of clusters (see fig. 5, col. 9, lines 1-20 and col. 13, lines 55-67 and col. 14, lines 1-8).

Therefore, it would have been obvious to a person of ordinary skill in the art at the time the invention was made to combine the teachings of Ioannidis with the teachings of Lundahl so as to have sum of squares of the result of the cluster to be compared and compare the multi-dimensional qualities data structure having multiple partitions and clusters that can be constructed for the same data. The motivation being to have a database for storing the data in which provides the clustering or partitioning advantage for a wider range of queries and used the bucket histogram technique to compare the distance of data distribution sets and the results from the clustering algorithms in a parallel processing system environment.

With respect to claims 5-6, Ioannidis discloses a method as discussed in claim 1. And Ioannidis teaches buckets (col. 9, lines 12-67).

Ioannidis teaches using bucket histogram technique for data clustering, the distance between of two multisets, in parallel processing database systems (col. 5, lines 25-50 and col. 6, lines 32-50), each bucket is assuming that the values that fall within the range of a bucket (col. 5, 52-67; also see col. 2, lines 16-65) and the frequencies representing sets of two data distributions (distance measurement based on various distribution moments) and comparing the data distribution sets. Ioannidis does not explicitly teach wherein the optimal number of clusters is determined by a maximum number of buckets for a variable, and wherein the optimal number of clusters is set to a threshold value in case the maximum number of buckets is greater than the threshold value.

However, Lundahl teaches the optimal of the number of clusters (see fig. 5, col. 9, lines 1-20 and col. 13, lines 55-67 and col. 14, lines 1-8) and the value of threshold (col. 13, lines 50-67, col. 21, lines 38-52 and col. 23, lines 36-61).

Therefore, it would have been obvious to a person of ordinary skill in the art at the time the invention was made to combine the teachings of Ioannidis in view of Martin with the teachings of Lundahl so as to have sum of squares of the result of the cluster to be compared and compare the multi-dimensional qualities data structure having multiple partitions and clusters that can be constructed for the same data (col. 4, lines 6-38, also see abstract and col. 3, lines 4-15). The motivation being to have a database for storing the data in which provides the clustering or partitioning advantage for a wider range of queries and used the bucket histogram technique to compare the distance of data distribution sets and the results from the clustering algorithms in a parallel processing system environment.

With respect to claims 7-9, Ioannidis discloses a method as discussed in claim 1. And Ioannidis teaches the relative foreground and background frequencies (during computing the distance of data distribution of bucket sets: col. 7, lines 15-67).

Ioannidis teaches using bucket histogram technique for data clustering, the distance between of two multisets, in parallel processing database systems (col. 5, lines 25-50 and col. 6, lines 32-50), each bucket is assuming that the values that fall within the range of a bucket (col. 5, 52-67; also see col. 2, lines 16-65) and the frequencies representing sets of two data distributions (distance measurement based on various distribution moments) and comparing the data distribution sets. Ioannidis does not

explicitly teach wherein the optimal number of clusters, a normalizing value and normalizing the result of the comparison, summing the results of the corresponding comparison values.

However, Lundahl teaches the optimal of the number of clusters (see fig. 5, col. 9, lines 1-20 and col. 13, lines 55-67 and col. 14, lines 1-8), multiplying and summing the result (the product of the matrix: col. 5, lines 24-51), and normalizing the values (col. 12, lines 27-67).

Therefore, it would have been obvious to a person of ordinary skill in the art at the time the invention was made to combine the teachings of Ioannidis with the teachings of Lundahl so as to have sum of squares of the result of the cluster to be compared and compare the multi-dimensional qualities data structure having multiple partitions and clusters that can be constructed for the same data (col. 4, lines 6-38, also see abstract and col. 3, lines 4-15). The motivation being to have a database for storing the data in which provides the clustering or partitioning advantage for a wider range of queries and used the bucket histogram technique to compare the distance of data distribution sets and the results from the clustering algorithms in a parallel processing system environment.

With respect to claim 10, Ioannidis teaches performing a number of data clustering operation (a number of operation to be applied on histogram technique for data distribution sets: col. 12, lines 4-67; also col. 4, lines 1-21).

Ioannidis teaches using bucket histogram technique for data clustering, the distance between of two multisets, in parallel processing database systems (col. 5, lines

25-50 and col. 6, lines 32-50), each bucket is assuming that the values that fall within the range of a bucket (col. 5, 52-67; also see col. 2, lines 16-65) and the frequencies representing sets of two data distributions (distance measurement based on various distribution moments) and comparing the data distribution sets. Ioannidis does not explicitly determining a quality index for each result of the data clustering operations; and c) selecting the result with the highest quality index as an end result of the data clustering.

However, Lundahl teaches the index and the best index value to be selected (col. 13, lines 8-67 and col. 14, lines 1-8) and the highest value to be chosen (col. 13, lines 30-42).

Therefore, it would have been obvious to a person of ordinary skill in the art at the time the invention was made to combine the teachings of Ioannidis with the teachings of Lundahl so as to have the best index value to be selected for the comparison and sum of squares of the result of the cluster to be compared and compare the multi-dimensional qualities data structure having multiple partitions and clusters that can be constructed for the same data (col. 4, lines 6-38, also see abstract and col. 3, lines 4-15). The motivation being to have a database for storing the data in which provides the clustering or partitioning advantage for a wider range of queries and used the bucket histogram technique to compare the distance of data distribution sets and the results from the clustering algorithms in a parallel processing system environment.

With respect to claim 11, Ioannidis teaches selecting an initial set of clusters (selecting a initial element in the bucket: col. 13, lines 1-58).

Ioannidis teaches using bucket histogram technique for data clustering, the distance between of two multisets, in parallel processing database systems (col. 5, lines 25-50 and col. 6, lines 32-50), each bucket is assuming that the values that fall within the range of a bucket (col. 5, 52-67; also see col. 2, lines 16-65) and the frequencies representing sets of two data distributions (distance measurement based on various distribution moments) and comparing the data distribution sets. Ioannidis does not explicitly teach determining a quality index for the clusters; and performing a number of iterations to improve the quality index.

x However, Lundahl teaches the number of iterations for the index to be chosen (col. 15, lines 1-26) and index and the best index value to be selected (col. 13, lines 8-67 and col. 14, lines 1-8).

Therefore, it would have been obvious to a person of ordinary skill in the art at the time the invention was made to combine the teachings of Ioannidis with the teachings of Lundahl so as to have the best index value to be selected for the comparison and sum of squares of the result of the cluster to be compared and compare the multi-dimensional qualities data structure having multiple partitions and clusters that can be constructed for the same data (col. 4, lines 6-38, also see abstract and col. 3, lines 4-15). The motivation being to have a database for storing the data in which provides the clustering or partitioning advantage for a wider range of queries and used the bucket histogram technique to compare the distance of data distribution sets

and the results from the clustering algorithms in a parallel processing system environment.

With respect to claim 12, Ioannidis teaches a method as discussed in claim 11.

Ioannidis teaches using bucket histogram technique for data clustering, the distance between of two multisets, in parallel processing database systems (col. 5, lines 25-50 and col. 6, lines 32-50), each bucket is assuming that the values that fall within the range of a bucket (col. 5, 52-67; also see col. 2, lines 16-65) and the frequencies representing sets of two data distributions (distance measurement based on various distribution moments) and comparing the data distribution sets. Ioannidis does not explicitly teach determining the quality index for the modified clusters, and using the modified clusters as a new initial set of clusters in case the quality index improved.

However, Lundahl teaches the index and the best index value to be selected (col. 13, lines 8-67 and col. 14, lines 1-8).

Therefore, it would have been obvious to a person of ordinary skill in the art at the time the invention was made to combine the teachings of Ioannidis with the teachings of Lundahl so as to have the best index value to be selected for the comparison and sum of squares of the result of the cluster to be compared and compare the multi-dimensional qualities data structure having multiple partitions and clusters that can be constructed for the same data (col. 4, lines 6-38, also see abstract and col. 3, lines 4-15). The motivation being to have a database for storing the data in which provides the clustering or partitioning advantage for a wider range of queries and used the bucket histogram technique to compare the distance of data distribution sets

and the results from the clustering algorithms in a parallel processing system environment.

Claim 13 is essentially the same as claim 1 except that it is directed to a computer program product rather than a method (), and is rejected for the same reason as applied to the claim 1 hereinabove.

Contact Information

6. Any inquiry concerning this communication or earlier communications from the examiner should be directed to Anh Ly whose telephone number is 703 306-4527 or via E-Mail: ANH.LY@USPTO.GOV. The examiner can normally be reached on 7:30 AM - 4:00 PM.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, John Breene, can be reached on 703 305-9790. The fax phone number for the organization where this application or proceeding is assigned is 703 746-7239.

Any response to this action should be mailed to:

Commissioner of Patents and Trademarks

Washington, D.C. 20231

or faxed to: Central Office (703) 872-9306 (Central Official Fax Number)

Hand-delivered responses should be brought to Crystal Park II, 2121 Crystal Drive, Arlington, VA, Fourth Floor (receptionist).

Any inquiry of a general nature or relating to the status of this application or proceeding should be directed to the receptionist whose telephone number is 703 308-6606 or 703 305-3900.

JEAN M. CORRIELUS
PRIMARY EXAMINER

ANH LY
MAY 19th, 2004